Routledge
Taylor & Francis Group

Journal of Media Practice

ISSN: 1468-2753 (Print) 2040-0926 (Online) Journal homepage: https://www.tandfonline.com/loi/rjmp20

Sounds too true to be good: diegetic infidelity – the case for sound in virtual reality

Angela McArthur, Rebecca Stewart & Mark Sandler

To cite this article: Angela McArthur, Rebecca Stewart & Mark Sandler (2017) Sounds too true to be good: diegetic infidelity – the case for sound in virtual reality, Journal of Media Practice, 18:1, 26-40, DOI: 10.1080/14682753.2017.1305840

To link to this article: https://doi.org/10.1080/14682753.2017.1305840

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 19 May 2017.

Submit your article to this journal

Article views: 1302

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rjmp21

JOURNAL OF MEDIA PRACTICE, 2017
VOL. 18, NO. 1, 26–40
http://dx.doi.org/10.1080/14682753.2017.1305840

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Sounds too true to be good: diegetic infidelity – the case for sound in virtual reality

Angela McArthur, Rebecca Stewart and Mark Sandler

Centre for Digital Music, Queen Mary University of London, UK

ABSTRACT

Cinematic virtual reality (VR) elicits new possibilities for the treatment of sound in space. Distinct from screen-based practices of filmmaking, diegetic sound–image relations in *immersive* environments present unique, potent affordances, in which content is at once imaginary, and real. However, a reductive modelling of environmental realism, in the name of 'presence' predominates. Yet cross-modal perception is a noisy, flickering representation of worlds. Treating our perceptual apparatus as stable, objective transducers, ignores the inter-subjective potential at the heart of immersive work, and situates users as passive spectators. This condescends to audiences and discounts the historic symbiosis of sound–image signification, which comes to constitute notions of verisimilitude. We understand the tropes; we willingly suspend disbelief. This article examines spatial sound rendering in virtual environments, probing at diegetic realism. It calls for an experimental, aesthetic approach, suggesting several speculative strategies, drawing from theories of embodied cognition and acousmatic practice (amongst others) which necessarily deal with space and time as contingencies of the immersive. VR affords a development of the dialectic between sound and image which distinctively involves our spatial attention. The lines between referent and signified blur; the mediation between representations invoked by practitioners, and those experienced by audiences, suggest new opportunities for co-authorship.

ARTICLE HISTORY
Received 16 January 2017
Accepted 9 March 2017

KEYWORDS
Spatial sound; virtual reality; aesthetics; realism; diegetic sound; immersion; perception

Introduction: spectacle of mes-mmersion

The growth in demand and supply for immersive experiences (Gartner 2016) whether for brand-building, entertainment, or empathic journalism, seems set to continue. The enablement of such experiences through new media such as virtual reality (VR) compounds what seems to be a desire for amplified experience. A posthuman encounter with the sublime, immersion in virtual environments allows temporary oblivion, an escape from real environments and the pressures-as-habitus of the information age. This in turn attests to the accessibility and convergence of technologies, rich, hyperreal media, and commerce's backing of this form of twenty-first century nihilism (Bia 2016).

Immersive experiences such as VR, are increasingly extensive (accommodating many sensory systems), surrounding (they approach us from any direction), vivid (rich and

CONTACT Angela McArthur a.mcarthur@qmul.ac.uk

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

varied in informational content) and match proprioceptive feedback about our body movements (Slater et al. 1996). As such, they dissolve subject/object oppositions, offering us complex multi-sensory environments in which our perceptual apparatus and technology are synthesised, with minimal mediation. VR is not in fact a new medium, but its level of current accessibility, brought about by improvements in computing, reduction in component size and reduced pricing, has made it a technology for the consumer. Yet it suffers from a lack of research due to the recency of its adoption (Biocca 1999; Sharples 2016) and exists in a fragmented ecosystem. Sound's late positioning in project timelines, often results in a pressured, innocuous approach to spatial sound design, just when it can be placed in *any* position relative to the user, and importantly, relative to its visual counterpoint – its source.

Production conventions for VR are underdeveloped compared with screen-based media, and evolve apace, assimilating the results of research and practice in a kind of internal combustion and accelerated reproduction. Components are developed and brought, 'market fresh' from high tech industrial zones in Schenzen. Cross-disciplinary, experimental work by engineers and psychologists working with virtual technologies continually redefine our conceptual frameworks for understanding the relationship between the hardware and human 'wetware',[1] work that is in turn absorbed into the pastiche-belly of wider dissemination: interactive VR installations, workshops held in cultural institutions, business meetups. Such empirically conducted research raises as many questions as it does provide affirmations or refutations, in a dense quagmire of ethical issues attempting to guide responsible use, for the young, for parents, for risk-averse though fervent manufacturers.[2] Will production and consumption guidelines emerge from trial and error?

Today's emphatic declamations ('thou shalt not use jump-cuts!') morph into tomorrow's anecdotes of derision. At present many challenges remain, from directing attention in a medium which necessarily deals with freedom of head movement (there is no finite framing as with screen-based media), to addressing emotional engagement and user interactivity (currently positioned as mutually exclusive). Are these obstacles to be overcome, or openings to new ways of conceiving communication? 'Hidden environments' (McLuhan and Fiore 1967; McLuhan and McLuhan 2011) – the subliminal forces of media and their effects – enable 'anti-environments' after all, the forms of work that rise to challenge and interrogate new media, and in doing so give rise to nascent creative strategies.

Some strategies will establish roots, over time bearing answers to VR's predicaments. They will form its material and political contingencies, and give VR meaning beyond affect. Presently, strategies of both environment and anti-environment are in flux. Meanwhile, an explicit recognition of the extent to which sound rendering in VR is influenced by its technologies, and their historicised ascent is useful.[3]

Convergence in technologies has placed production into the hands of novice practitioners, and reception into the ears of naïve listeners, using (minimally) a smart phone, cardboard headset, and their own headphones. Binaural sound synthesis, dynamically rendered, offers a compelling experience even under such circumstances. In pursuit of 'presence' – a sense of 'being there', improved realism and quality may seem of key concern, indeed they exert a powerful effect (Hendrix and Barfield 1996; Serafin and Serafin 2004). Yet does ever-increasing fidelity produces ever-increasing aesthetic gains? We may feel we are 'there', but once the impact of VR's novelty has lessened, why is 'there' interesting? What else could sound contribute?

Audio in VR has distinct benefits, including its ability to cue user attention. It can be, however, the victim of its own ephemeral appeal. For screen-based filmmaking, sound has been a post-production concern. Automated dialogue replacement, foley, and sound effects have all contributed to its efficacy in 'adding value' to the visual in a process of 'synchresis', the forging of sound–image signification (Chion 1994). Sound adds value by creating:

> … the definite impression, in the immediate or remembered experience one has of it, that this information or expression 'naturally' comes from what is seen, and is already contained in the image itself. Added value is what gives the (eminently incorrect) impression that sound is unnecessary, that sound merely duplicates a meaning which in reality it brings about, either all on its own or by discrepancies between it and the image. (Chion 1994, 5)

Chion draws our attention to the crucial role sound plays in creating, legitimising and subverting meaning. Yet it remains invisible and consequently to some, even if involved in production, imperceptible. Acknowledged more in absence than presence.

In VR, post-production is bottlenecked in the extreme, exposing for the first time the impact of earlier mistakes, often unrecoverable. New technologies require new skills, new constellations of teams, familiarity with products and processes that rapidly shift. Many practitioners learn through failure. Sound – if making its inception here – precariously remains a post-production affair, *despite* the extent of its contribution. At present diegetic sound is usually mapped to its visual source, whilst non-diegetic sound is placed 'in head' stereophonically. This is not only a result of time constraints, however. Unlike a visual 'shot' sound in any media does not exist as a discrete unit, thus lacking the '… enormous advantage of being a neutral unit, objectively defined, that everyone who has made the film as well as those who watch it can agree on' (Chion 1994, 41).

Can audiophiles be forgiven similar transgressions and visual biases? The '… discourse of loss' in which audiophiles are typically obsessed with fidelity and reproduction of the live original as the technological ideal …'. (Chow and Steintrager 2011, 5) suggests that even if sound is acknowledged, it represents almost a fetish in absentia. Chow and Steintrager see both sound's technical and phenomenological pursuit as predicated on its elusive qualities, its unwillingness to be caught – ' … it has always already escaped' (2011, 5). How faithfully then, was it ever captured?

According to Altman (1992), recordings have only partial correspondence to the original event. Chion underlines the reasoning underlying this approach in film sound:

> … the processed food of location sounds is most often skimmed of certain substances and enriched with others. Can we hear a great ecological cry – 'give us organic sound without additives?' Occasionally filmmakers have tried this […] the result is totally strange. Is this because the spectator isn't accustomed to it? Surely. But also because reality is one thing, and its transposition into audiovisual two-dimensionality […] which involves radical sensory reduction, is another. (1994, 96)

Chion perhaps stops short where he could challenge the 'realism' of perception itself.[4] The notion of veridicality between the world and our perception of it, is questionable. It presupposes that our sensory apparatus are incidental, 'faithful transducers' replicating objective external reality (Lennox 2004). Such ideals of objectivity and

> … presumptions of universality have […] led scholars to treat sounds as stable objects that have predictable, often technologically determined, effects on a generalised perceptual

> consciousness [… which] did not explicitly recognise the constitutive differences that participate in the 'soundscape' as a multivalent field of sounds with diverging social identities, individual creativities and affordances, biodiversities and differing abilities. (Novak and Sakakeeny 2015, 7)

The world 'out there' is not objectively experienced by us then. It – like media sound – is always representational. Chion goes on:

> … conventions of rendering, sound effects, and so forth, […] consist of accommodations and adjustments […] to conserve a certain sense of realism and truth in their new representational context. […] 'trompe-l'oreille' is a worthy art … . (1994, 96)

Rendering is a worthy act of creation then, it declaims a voice, a message, an identity. It does not purport to be non-partisan. Our attraction to rendered hyperreality, immersion, and other rich environments testifies: Unreality is more presence-inducing than reality. So then, realism need not be the aim. What would Chion say about virtual, rather than two-dimensional media? At what attentional cost do we enter immersive unrealism?

Unbounded by frame, we have more information to parse in VR. Our spatial attention must be allocated and coordinated across modalities, each of which initially encode information differently. Perceptual processing and prioritisation of information from different modalities involves complex interactions. Given our limited attentional resource,[5] we may favour 'reliable' sensory input over accuracy (Battaglia, Jacobs, and Aslin 2003). How is it that ultimately we are presented with a coherent perceptual experience? How much are we subjectively 'gathering in' loose ends?

> This work of gathering – an effort to unify and make cohere – implies that subjectivity is involved whenever we try to draw some boundary in the sonic domain … . (Chow and Steintrager 2011, 2)

There are adaptive benefits to the resolution of conflicting sensory inputs, benefits that help explain the 'Ventriloquist Effect' (Alais and Burr 2004). The question then arises of how we *enjoy* experiences if they rupture such correspondence, and require a higher level of attentional resourcing?

> It would appear that the detection of an irregular event is a prerequisite for an emotional response. (Steinbeis, Koelsch, and Sloboda 2006, 1391)

Is such irregularity that which violates our expectations, jolts our complacency, and ultimately moves us? Perhaps the higher cognitive load which such a situation demands could be used to transport users along an aesthetic arc? Perhaps a one-to-one diegetic mapping of sound to image isn't the most effective way of communicating information, particularly in an artistic context? Drawing from music cognition studies we might explore the relationship between perception and aesthetics. In the inverted bell-curve of the complexity-liking relationship (North and Hargreaves 1995; Orr and Ohlsson 2005), content too predictable or too complex is less 'liked', whilst at the peak of the inverted bell, complexity and liking are in an optimal position. Might sound–image incongruence, initially unexpected yet systematically applied (thus recognisable), move users along a similar aesthetic arc? What might the optimal position be for liking/complexity? Would this vary between audiences, or for the same person over time? Could it be exploited to sustain interest?

> It is not the smooth simulacrum of reality but the ambiguity in the continually shifting background of sound-image relations that serves to hold the viewer-auditor's interest, promising suspense and surprise even where […] there is no narrative. (Johnson 1989, 26)

Complexity – the *impression* of randomness despite regularity in incongruent sound–image relations, can 'continually shift', can be increased or decreased to induce a heterogeneous, emergent experience. The site of this 'becoming' experience is distributed, with cognitive processes acting to organise further meaning from arranged audio-visual phenomena. Cognitive processes themselves contain personal and collective histories, such that the act of reception is also one of authorship. There is no supplicant with abdicated sensoria. VR is socially composed with a level of inter-subjective production which may be unprecedented – individual differences are exacerbated in immersive settings due to the egocentric placement of audiences, and the orbiting of events and constituents such that 'everything is interpreted relative to you' (Blesser and Salter 2009, 49).

What understanding then, need practitioners of spatial sound design for immersive media have, of cognition? As an example, Chion discusses a kind of substitute memory:

> The question of verisimilitude, is a terribly ambiguous and complicated one […] sound that rings true for the spectator and sound that is true are two very different things. In order to assess the truth of a sound, we refer much more to codes established by cinema itself, by television, and narrative representational arts in general […] quite often we have no personal memory we might refer to regarding a scene we see. (1994, 107)

Here media acts as a substitute, codified recollection, not just due to a lack of experience, but as Chion explains, also resulting from the strength of impression left by media. This in turn focuses our concern on sound's *rendering* more than its reproduction, even in a reality that is virtual. This symbiosis, and the partiality of our auditory 'representational' archives, has been acknowledged:

> Whereas comparing the visual architecture of two spaces through pictures does not place a burden on short-term memory, comparing the aural architecture of two spaces involves […] the unreliability of auditory memory … . (Blesser and Salter 2009, 17)

Despite Altman's (1992) 'reproductive' fallacy in cinema sound (which holds that the image is creatively unfaithful, but sound is automatically faithful), we can see that practitioners have exploited sound–image disjuncture and recognised audience cognitive subjectivities before immersive media assumed its current forms. The line between diegesis and mimesis was blurred by film and theatre sound–image relations, to effect. Some of these effects may be so well established that they exist below the threshold of awareness. A banal example – in real life we would not hear a cat purring unless we were extremely close to it. In film sound practice an incredulously amplified purr is an acceptable representation of reality.[6] This acceptance underlines the value of expression over realism. Tarkovsky clearly states the case:

> … sounds of the world reproduced naturalistically in cinema are impossible to imagine: there would be a cacophony. Everything that appeared on the screen would have to be heard on the soundtrack, and the result would amount to sound not being treated at all in film. If there is no selection then the film is tantamount to silent, since it has no sound expression of its own. (1989, 161)

When dealing with sound that has been designed, we *necessarily* do so aesthetically (aesthetic appreciation as perceptual experience). Yet our symbiotic absorption of mediated 'realities' means we soon incorporate them into an updated range of realism. The treatment of sound–image relations may want its aesthetic expression, but we could exercise caution here. A critique of the hyperreal as a semiotic structure and nihilistic device provides a useful reminder (Eco 1986, 1989; Böhme 1993; Baudrillard, 1994) that doppelgängers are elusive figures whom we would do well to recognise, less they succeed as the signified, and flatten our sense of perspective. As Böhme prompts, any 'ordering' of elements contains some agenda, whether declared or implicit. Without interrogation, we risk perpetuating hierarchies and inequalities in the most benign or beguiling of forms:

> [A] response to the progressive aestheticisation of reality … aesthetics represents a real social power. There are aesthetic needs and an aesthetic supply … to the aesthetics of the work of art we can now add … the aesthetics of everyday life, the aesthetics of commodities and a political aesthetics. (Böhme 1993, 125)

Our complicity with spectacle neuters us (at least in part). Thus there is a certain morality in imagination, and the ability to draw ourselves away from an enveloping, mes-mmersive present toward disjuncture, the rupturing of expectation, the artwork as 'anti-environment'.

Diegetic anti-realism

> The real can never be represented; representation alone can be represented. For in order to be represented, the real must be known, and knowledge is always already a form of representation. (Altman 1992, 46)

Realism, and so anti-realism, is a continually shifting concern. VR's 'reality' is bound by inherited representational systems until its own emerge. And whilst they do, we might ask – what affordances does VR offer for effective communication? Tarkovsky asserts:

> … accurately recorded sound adds nothing to the image system of cinema, for it still has no aesthetic content […] if the real sounds are distorted so that they no longer correspond with the image – then the film acquires a resonance. (1989, 162)

Such deviations, applied to three dimensions in a virtual experience, may engender effective strategies for designing spatial sound in a world that is wilfully incomplete, awaiting the interpretive 'filling up' by individual users' meaning.

> We need an alternative concept that assumes neither the completeness nor the consistency of a real space. (Blesser and Salter 2009, 132)

Exaggerated sound–image interactions may for example 'break' the representational frame in an otherwise unframed environment, allowing for real-time subjectivity and awareness of illusion, an ebb and flow of tension and aesthetic release. This may all be achievable in an environment which promotes phenomenological presence. But rule-breaking is an act involving less deliberation than rule-making:

> When inventing new rules and applying them in novel ways, an artist is just as likely to create musical experiments that have little enduring value. The application of aural architecture to cinema is a good example of aesthetically pleasing spatial rules that never presume a

> space as a real environment. […] unrelated aural and visual spaces often coexist simultaneously … . (Blesser and Salter 2009, 160)

Rupturing for the sake of rupturing will only take us so far; too much may break presence and fail to engage audiences (who have an overwhelming array of choice and lack of time). Too little may go unnoticed. Experimentation is requisite, and it is contended that a systematic, considered strategy will be less likely to become the very thing it seeks to subvert – spectacle.

Speculative strategies for experimentation

> … the history of music illustrates the attempt to find ways of describing, notating, and therefore identifying sounds, without specifying a cause for them … . (Scruton 1997, 3)

From whom should we draw inspiration for spatial sound design? An experimental palette may include a diverse a range of disciplines (a hybrid approach being appropriate for this hybrid medium).

Aesthetic philosophy

Aesthetic philosophy, working towards a new aesthetic of 'acoustic atmospheres' (Böhme 1993) would allow us to consider the relationship between environmental qualities and human states. Böhme suggests we treat aesthetics experientially rather than dialectically. This suits VR as immersive environment:

> This new aesthetics circumscribes […] their very immersion […] Atmosphere surrounds, includes, involves, envelops, and gives forth both the qualities of the environment and the experiencing human […] in order to stand as atmosphere, a spatial arrangement needs to be experienced or imagined into being. (Grant 2014, 21)

But should we rely on its sensuous or formal characteristics without concern for the social function of the work? Böhme cautioned against this. Updating metaphors of materiality in filmmaking practice (for example those of the Constructivists) requires that we celebrate the innate sensual properties of aesthetics in spatial sound design, and the dissolution of subject–object authorship. Here we may look to theories of embodied cognition.

Embodied cognition

Theories of embodied cognition suggest that form and meaning cannot be phenomenally or logically separated.[7] Perception directly engages with meanings that are in the world and body, but uses symbolic representational structures to do so. This is a useful segue into a strategy which leverages the phenomenology of experience whilst (through designed-in incongruence), foregrounding our subjectivities.

Lakoff and Johnson (2003) offer conceptual metaphors and image schemas[8] for understanding the way that we relate to sound (a more abstract phenomena) in terms of our concrete experience. This process in turn leads to systematic conceptual metaphors and related expressions, and might provide a useful framework for incongruence. As an example strategy, incongruence might deviate aesthetically by slowing the tempo of sounds rising in pitch and quickening the tempo of those falling. This would represent

a clear disjuncture, whilst adhering to principles of concrete experience (we slow down when moving upwards due to the greater exertion needed to counter gravity's effect). This might provide artistic interest whilst not 'breaking' presence.

Embodiment is not, however, de facto reflexive. In discussing Benjamin's (1936) essay on the effect of reproduction on the 'aura' of a work of art, Böhme argues that the avant garde:

> … did not succeed in discarding aura like a coat […] What they did succeed in doing was to thematize the aura […] this made it clear that what makes a work an artwork cannot be grasped soley through its concrete qualities. (1993, 116–117)

He goes on to cite Benjamin's description of aura as spatio-temporally contingent, and apt as analogous to an atmosphere, internalised through breathing, entering the 'bodily economy of tension and expansion' (Böhme 1993, 117) in an act of inter-subjective absorption, extension, and spatial perception. Layering, at the concrete and conceptual strata, means its

> … human character is perceivable […] audiences are capable of perceiving properties of works as realizing artists' manipulations of materials … . (Gerwen 2012, 223)

Crediting audiences with this ability echoes Chion's (1999) concept of the 'acousmêtre' that calls attention to the false sense of unity in sound–image relations, provided through synchrony. The acousmêtre, being ambiguous, leaves the source of the sound open to interpretation; it destabilises conditions. The anti-environment continually shifts. Which practical examples can we draw from, which tackle sound in space phenomenologically and conceptually? Sound studies critique sound beyond both its formal and ideological characteristics, even those which seem nebulous, such as its ubiquity and panopticism. This bodes well for immersion.

Sound studies and acousmatic practice

Recent developments in sound studies are encouraging. 'Acoustemology' (acoustics + epistemology) for example, theorises sound:

> … as a way of knowing. In doing so it inquires into what is knowable, through sound and listening. (Feld 2015, 12)

This kind of approach necessarily deals with socially constructed knowledge and practice, refusing sound's 'objective universality' by positioning itself:

> … against 'soundscape', the key legacy term associated with [R Murray] Schafer [acoustemology] refuses to sonically analogise or appropriate 'landscape', with all its physical distance from agency and perception. (Feld 2015, 15)

Meanwhile scholars and practitioners of acousmatic and electroacoustic music deal with notions of space as essential components. Smalley's (1997) notions of source-bonding or surrogacy are two foundations for compositional strategies. Source bonding can be real or imagined and is:

> The natural tendency to relate sounds to supposed sources and causes, and to relate sounds to each other because they appear to have shared or associated origins. (1997, 110)

Surrogacy meanwhile describes the progressive remoteness of sound source from directly experienced physical gestures. Smalley (1997, 124) provides guidelines for a global *spatial style* for a work[9] which could be applied to spatial sound design for VR (indeed his first guideline, that of 'single spatial setting', a cumulative spatial awareness in which a user appreciates (over time) the topology of a space within which a work sits, can be seen at work in 'Stifled' – a VR game released in 2017 where the world is revealed through the sound-input of a user's microphone).[10]

In his discussion of such ideas, Emmerson discusses how our auditory processing utilises:

> … established frames of reference when confronted with spaces, real or imaginary. Using electronics we may conjure up this increasingly wide set of alternatives. On the one hand the composer may attempt spatial re-presentation, that is the re-creation of an appropriate 'real' space to support a narrative; this can approach a kind of onomatopoeia in which the idiosyncrasies of a real space maybe mimicked … we may move from direct imitation […] through increasingly vague evocations to more remote impressions of colour and texture. (2007, 101)

He describes this range of options as exactly parallel to the mimetic axis of the materials of electroacoustic music. Gerwen's (2012) caution that we cannot hear everyday sounds acousmatically, as we cannot abstract sound events from their objects (including our own bodies), is a useful caveat. Sound may be intentionally organised and decoded, but its 'sounding' (its environmental and performative realisation) is integral to its appreciation as aesthetic expression. Gerwen, though not the first to do so, calls for critical listening.

Classical cognitivism

A classical cognitivist view may be useful to understand aesthetics in the context of expectation. Composer Frank Lerdahl usefully states:

> Aesthetic Claim 1: The best music utilizes the full potential of our cognitive resources.
>
> Aesthetic Claim 2: The best music arises from an alliance of a compositional grammar with the listening grammar. (1992, 119)

Lerdahl's definition of musical grammar, is 'a limited set of rules that can generate indefinitely large sets of musical events and/or their structural descriptions' (1992, 99). He subdivides this into two elements – compositional grammar, consciously employed to generate and organise events, and listening grammar, more or less unconsciously employed by auditors, and in effect generating mental representations of the music.

Lerdahl points to the prerequisite for 'stability conditions' to achieve comprehension and avoid boredom in the experience of listening (expectation and the violation of expectation combine to form satisfying musical experience where listener is neither constantly surprised nor constantly correct in their predictions). Transitional probabilities – the statistical regularities of sequences of events – can be learnt with surprisingly little exposure, with semantically impoverished stimuli, and by infants (Saffron et al. 1999). The learning and exploitation of these stability conditions are engendered by systematically applied incongruence, which can then be ruptured to destabilise for aesthetic or narrative interest. Sound may be exceptionally placed to respond to the ideology of stability, which according to Voegelin:

> … does not exist but is assumed and pretended by a visual ideology. Sound by contrast negates stability through the force of sensory experience … . (2010, 11–12)

We cannot exert control over sound's ability to move us, we cannot shut it out. Audition is the eternal 'site of performative embodiment' (Barton and Windeyer 2012, 198) which extends beyond our will, enacting our bodies as environmental alarms. Sound is not stable, nor as we have seen, is our perception of it. How then, to allow for this in our ways of discussing it? How to mirror this fluidity in the very syntax with which we describe audio-vision?

Taxonomies

A final strategy, that of re-thinking taxonomies of sound–image, is now suggested. Many have contributed to establishing classifications in this field (Percheron and Butzel 1980; Raskin 1992; Chion 1994) to address our visually biased syntax for film (as Altman (1992) points out, we go to 'see' a film, we discuss 'point of *view*' etc.).

> Even now the cinema has kept its ontologically visual definition no less intact. A film without sound remains a film; a film with no image, or at least without a visual frame for projection, is not a film. Except conceptually. (Chion 1994, 143)

This biasing persists despite sound (unlike image) retaining its dimensionality in media. Image, whether screen-based or post-screen, is flattened. It flickers. We see stitch-lines, pixilation and other artefacts of its reproduction:

> … a filmed object loses a dimension in the recording, recorded sound maintains its dimensions … . Recorded sound thus has a higher coefficient of 'reality' than the image … . (Stam 2000, 214)

Addressing this insidious imbalance may require new schema. Raskin (1992) situates both actual and subjective sound within the boundaries of the diegetic, seeing sound as a typology:

1) as a means for including in a stylistic profile of a given director, an exact description of his/her sound 'palette' - perhaps even as it evolves from film to film;
2) as a construct enabling us to deal with clearly defined varieties of sound, one at a time, in a systematic effort to chart the functions of film sound; and
3) as a basis for determining to what degree any given model for studying film aesthetics, encompasses a full range of variables with respect to sound. (1992, 12)

The idea of 'incongruent' (or at least playful) taxonomies is best reported by Foucault. We can make such rich, idiosyncratic categorisation our own, and unbind it from historical, objective alterity.

> … animals are divided into: '(a) belonging to the Emperor, (b) embalmed, (c) tame, (d) suckling pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i) frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) et cetera, (m) having just broken the water pitcher, (n) that from a long way off look like flies'. In the wonderment of this taxonomy, the thing we apprehend […] as the exotic charm of another system of thought, is the limitation of our own, the stark impossibility of thinking that. (Foucault 1970, xv)

Such taxonomy may not relate directly to sound–image relations in VR, but the freedom and possibility underpinning them may rupture perceived wisdom, in favour of unperceived imagination.

Wrapping up: keeping it unreal

Without the sound technologies to enable dynamic spatial rendering, the question of keeping it real or unreal is moot. Intentional rupturing would not be possible, only unintentional (wild) glitching. Deliberate rupturing must *exceed* realism in some sense, and could usefully defer to the reception of media as an act of co-constitution:

> If we expect spatial music to be something that a listener can be inside and can explore, then it may be necessary to surrender some control to the listener. (Lennox 2004, 30)

Enabling audiences freedom to engage their own subjectivities offers spatial audio aesthetics as something which non-spatial audio cannot be. According to Lennox (2004) there cannot be a strong argument that the only kind of spatial structure we could understand is that found in real environments. We may use the proxy of immersive experience to (at least temporarily) discharge residual anxieties, a mental 'stress ball', absorbing the excesses of a digital existence with which our corporeality struggles. Yet we need to look beyond such techno-seductive affects. Immersive experiences may seem less mediated than ever, but the complex and multifarious systems they arise from and serve contain their own organisational aesthetics. Buyer beware.

New practices for the presentation of multi-sensory experience will emerge and ossify, borrowing from prior mediums as they do. Any claims of objective realism uphold hidden environments, and the move from real to virtual environment is uncanny. In designing sound for space, practitioners may need to do more than upskill technically, as they carry new ethical responsibilities forward. Yet the promise of this medium, of an aesthetic, immersive space in which sound is crucial and can be highly designed, and variously decoded, is incentive enough.

In evaluating the aesthetics of spatial audio in VR, we might recall Chion's (1994) consideration of Tati's work, and ask ourselves, as he does, whether the:

> … audiovisual strategy produce[s] added value? That is, are we dealing with sound that enlivens the image, and deepens it in spatial terms? (1994, 125)

Notes

1. Two useful authors to review are Mel Slater, Research Professor at the University of Barcelona whose work uses VR to examine body ownership illusions as studied in cognitive neuroscience, and Pontus Larsson, whose research encompasses multimodal interaction, presence and virtual acoustics, and who currently working with Volvo Technology within the Human Factors group often work in interdisciplinary research teams.
2. For a detailed account including recommendations for researchers and consumers, see Madary and Metzinger (2016) 'Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology'.
3. For an account of the historical development of sound in early cinema which foregrounds the triumph of idiosyncratic practices over rationally presented guidance, see Altman's essay 'Sound Space' in Altman (1992).

4. As examples, Bregman's (1990) 'Auditory Scene Analysis: The perceptual Organization of Sound', and Spence's Crossmodal Research Lab at Oxford University, alert us to the perceptual illusions which underlie our reasonable if flawed impressions of the sensory world.
5. The concept of human attention (behavioural and cognitive processing) as a limited resource which can be allocated by concentrating certain aspects of information while ignoring other others.
6. For a discussion of these concerns see Bottomore's (1999; 2001) work on sound practices in early cinema.
7. For an introduction see Anderson (2003) 'Embodied Cognition: A Field Guide' and Wilson (2002) 'Six Views of Embodied Cognition'.
8. Lakoff and Johnson 'Metaphors We Live By' (2003) 2nd ed.; Johnson (1987), 'The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason'; Lakoff and Johnson (1999), 'Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought'. Lakoff and Johnson argue that (particularly metaphorical) language is not simply a mental construction, but is shaped by bodily mediation of the world, and so our corporeal experience shapes our thinking, and is not distinct from cognition.
9. (1) Single spatial setting. The single setting has two aspects. A work can be set in a single type of space of which the listener is aware at the outset. On the other hand, different aspects of a space can be revealed over time. Spatial awareness is cumulative, and the listener eventually realises that there is a global spatial topology into which the whole work fits. For example, the extremes of proximity and distance are unlikely to be known until the work has advanced somewhat. (2) Multiple spatial settings. Throughout the work, the listener is aware of different types of space which cannot be resolved into a single setting. (3) Spatial simultaneity. Imagine a very present granular texture directly in front of you as if actually within your listening space, while in the distance a door closes in a large reverberant space. You are aware of simultaneous spaces. (4) Implied spatial simultaneity. Implied simultaneity occurs when the listener remains aware of the existence of a space in its absence. This can occur, for example, when contrasting spaces are intercut and alternated (spatial interpolation), giving the impression of simultaneity even though the spaces are presented successively. This is related to film, where in spite of the cutting between successive events, they are considered concurrent. (5) Spatial passage. Passage between spaces can be sudden (interrupted passage), repeatedly intercut (interpolated passage) or more gradually merged (graduated passage). (6) Spatial equilibrium. What is the relative balance between types of perspective and spatial texture in the work? Is one type of space emphasised more than another? Are there alternations or reciprocal exchanges between spaces?
10. See http://store.steampowered.com/app/514830.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors wish to acknowledge support from the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology through Queen Mary University of London.

Notes on contributors

Angela McArthur is a practising artist and PhD student on the Media, Arts & Tech programme at Queen Mary University London. Her work explores spatial sound for immersive environments and virtual reality, synthesising technical and creative practices with auditory perception research. She

has worked with the BBC R&D lab, and is currently in post production for an experimental sound-led cinematic VR film, which aims to embody and further develop her conceptual framework.

Dr Rebecca Stewart is a Lecturer in the School of Electronic Engineering and Computer Science at Queen Mary University of London. She works with e-textiles and signal processing to build interactive, body-centric wearable computing systems which often incorporate performance, fashion, music and/or design. As a member of the Centre for Digital Music and the Centre for Intelligent Sensing, she also conducts research into binaural audio for creative applications.

Prof Mark Sandler, FREng, FAES, FIET, FIEEE, CEng, is Founding Director of the Centre for Digital Music, a world-leading research group in audio and music technology with over 80 members. The Centre is in Queen Mary University of London's School of Electronic Engineering & Computer Science, where he holds the chair in Signal Processing. He is Principal Investigator of the EPSRC-funded grant, Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption (www.semanticaudio.ac.uk). He is a recipient of the Royal Society Wolfson Research Merit Award (2015–2019) and has published over 400 papers in conferences and journals.

ORCID

Angela McArthur http://orcid.org/0000-0002-4564-4737
Rebecca Stewart http://orcid.org/0000-0001-6976-8024
Mark Sandler http://orcid.org/0000-0002-5691-8107

References

Alais, D., and D. Burr. 2004. "The Ventriloquist Effect Results from Near-Optimal Bimodal Integration." *Current Biology* 14: 257–262.

Altman, R., ed. 1992. *Sound Theory, Sound Practice*. New York: Routledge.

Anderson, M. L. 2003. "Embodied Cognition: A Field Guide." *Artificial Intelligence* 149 (1): 91–130.

Barton, B., and R. Windeyer. 2012. "Binaural Perspectives on Site-Specific Sound." In *Performing Site-Specific Theatre: Politics, Place, Practice*, edited by A Birch and J Tompkins (pp. 182–200). London: Palgrave Macmillan.

Battaglia, Peter W., Robert A. Jacobs, and Richard N. Aslin. 2003. "Bayesian Integration of Visual and Auditory Signals for Spatial Localization." *Journal of the Optical Society of America A* 20: 1391–1397.

Baudrillard, J. 1994. *Simulacra and Simulation*. Ann Arbor: University of Michigan Press.

Benjamin, W. 1936. *The Work of Art in the Age of Mechanical Reproduction.* London and New York: Penguin Books.

Bia, Bord. 2016. "Immersive Experiences." Accessed February 6, 2017. http://www.bordbiaconsumerlifestyletrends.ie/3-shared-experiences-sub-trend-immersive-experiences/.

Biocca, F. 1999. "Lab Probes Long-term Effects of Exposure to Virtual Reality", Interview via Johnson, C for EE Times. Accessed December 7, 2016. http://www.eetimes.com/document.asp?doc_id=1138729.

Blesser, B., and L.-R. Salter. 2009. *Spaces Speak, Are You Listening? Experiencing Aural Architecture*. Cambridge, MA: MIT Press.

Bregman, A. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA, US: The MIT Press.

Böhme, G. 1993. "Atmosphere as the Fundamental Concept of a New Aesthetics." *Thesis Eleven* 36 (1): 113–126.

Bottomore, S. 1999. "An International Survey of Sound Effects in Early Cinema." *Film History* 11 (4): 485–498.

Bottomore, S. 2001. "The Story of Percy Peashaker: Debates about Sound Effects in the Early Cinema." In *The Sounds of Early Cinema*, edited by Richard Abel and Rick Altman (pp. 129–142). Bloomington and Indianapolis: Indiana University Press.

Chion, M. 1994. *Audio-Vision: Sound on Screen*. Translated by Claudia Gorbman. New York: Columbia University Press.
Chion, M. 1999. *The Voice in Cinema*. Translated by Claudia Gorbman. New York: Columbia University Press.
Chow, R., and J. Steintrager. 2011. "In Pursuit of the Object of Sound: An Introduction." *Differences: The Sense of Sound* 22: 2–3.
Eco, U. 1986. *Faith in Fakes: Travels in Hyperreality*. London: Vintage/Random House.
Eco, U. 1989. *The Open Work*. Cambridge, MA: Harvard University Press.
Emmerson, S. 2007. *Living Electronic Music*. Farnham, UK: Ashgate.
Feld, S. 2015. "Acoustemology." In *Keywords in Sound*, edited by D Novak and M Sakakeeny (pp. 12–21). Durham and London: Duke Uni Press.
Foucault, M. 1970. *The Order of Things*. New York: Pantheon.
Gartner. 2016. "Hype Cycle for Emerging Technologies, 2016." Accessed February 6, 2017. https://www.gartner.com/doc/3383817.
Gerwen, R. V. 2012. "Hearing Musicians Making Music: A Critique of Roger Scruton on Acousmatic Experience." *The Journal of Aesthetics and Art Criticism* 70 (2): 223–230.
Grant, S. 2014. "Performing an Aesthetics of Atmospheres." *Literature & Aesthetics* 23 (1): 12–32.
Hendrix, C., and W. Barfield. 1996. "The Sense of Presence within Auditory Virtual Environments." *Teleoperators & Virtual Environments* 5: 290–301.
Johnson, M. 1987. *The Body in the Mind: The Bodily Basis of Imagination, Reason, and Meaning*. Chicago: University of Chicago Press.
Johnson, W. 1989. "Sound and Image: A Further Hearing." *Film Quarterly* 43 (1): 24–35.
Lakoff, G., and M. Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western thought*. New York: Basic Books.
Lakoff, G., and M. Johnson. 2003. *Metaphors we Live by*. London: The University of Chicago Press.
Lennox, P. 2004. "The Philosophy of Perception in Artificial Auditory Environments: Spatial Sound and Music." PhD thesis, University of York.
Lerdahl, F. 1992. "Cognitive Constraints on Compositional Systems." In *Contemporary Music Review*. Vol. 6, 97–121. Glasgow: Harwood Academic.
Madary, M., and T. Metzinger. 2016. *Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology* in Front. Robot. AI 3:3.
McLuhan, M., and Q. Fiore. 1967. *The Medium is the Message: An Inventory of Effects*. New York: Bantam Books.
McLuhan, M., and E. McLuhan. 2011. *Media and Formal Cause*. Houston, TX: NeoPoiesis Press.
North, A., and David J. Hargreaves. 1995. "Subjective Complexity, Familiarity, and Liking for Popular Music." *Psychomusicology: A Journal of Research in Music Cognition* 14 (1–2): 77–93.
Novak, D., and M. Sakakeeny, eds. 2015. *Keywords in Sound*. Durham and London: Duke Uni Press.
Orr, M., and S. Ohlsson. 2005. "Relationship Between Complexity and Liking as a Function of Expertise." *Music Perception: An Interdisciplinary Journal* 22 (4): 583–611.
Percheron, D., and M. Butzel. 1980. "Sound in Cinema and its Relationship to Image and Diegesis." *Yale French Studies, Cinema/Sound* 60: 16–23.
Raskin, Richard. 1992. "Varieties of Film Sound: A New Typology." fra (Pré) Publications, Romansk Inst. Aarhus Universitet, April 1992.
Saffron, Jenny R., Elizabeth K. Johnson, Richard N. Aslin, and Elissa L. Newport. 1999. "Statistical Learning of Tone Sequences by Human Infants and Adults." *Cognition* 70: 27–52.
Scruton, R. 1997. *The Aesthetics of Music*. Oxford: Clarendon Press.
Serafin, S., and G. Serafin. 2004. "Sound Design to Enhance Presence in Photorealistic Virtual Reality." Proceedings of ICAD 04-Tenth meeting of the International Conference on Auditory Display, Sydney, Australia, July.
Sharples, S. 2016. *Long-term effects of virtual reality use need more research, say scientists*, interview via Davis, N for Guardian. Accessed December 7, 2016. https://www.theguardian.com/technology/2016/mar/19/long-term-effects-of-virtual-reality-use-need-more-research-say-scientists.
Slater, M., V. Linakis, M. Usoh, R. Kooper, and G. Street. 1996. "Immersion, Presence, and Performance in Virtual Environments: An Experiment with Tri-Dimensional Chess." In *Proceedings of ACM*

Symposium on Virtual Reality Software and Technology (VRST), (Vol. 163, p. 72) (pp. 163–172). New York, NY: ACM Press

Smalley, D. 1997. "Spectromorphology: Explaining Sound-Shapes." *Organised Sound* 2: 107–126.

Stam, R. 2000. *Film Theory: An Introduction*. New York: Blackwell.

Steinbeis, N., S. Koelsch, and J. A. Sloboda. 2006. "The Role of Harmonic Expectancy Violations in Musical Emotions: Evidence from Subjective, Physiological, and Neural Responses." *Journal of Cognitive Neuroscience* 18: 1380–1393.

Tarkovsky, A. 1989. *Sculpting in Time: Reflections on the Cinema*. New ed. Austin, TX: University of Texas Press.

Voegelin, S. 2010. *Listening to Noise and Silence: Towards a Philosophy of Sound Art*. New York and London: Bloomsbury.

Wilson, M. 2002. "Six Views of Embodied Cognition." *Psychonomic Bulletin & Review* 9 (4): 625–636.